Executive Steering Committee
For A.C.E. Policy II
(ESCAP II)
Report 18
October 11, 2001

ESCAP II: P-Sample Nonmatch Analysis

Glenn Wolfgang,
Tamara Adams,
Peter Davis,
Xijian Liu, and
Phawn Stallone
Decennial Statistical
Studies Division

U S C E N S U S B U R E A U
Helping You Make Informed Decisions

CONTENTS

LIST OF TABLES

EXECUTIVE SUMMARY

The goal of the P-Sample Nonmatch Analysis was to identify characteristics that may be related to being missed in census enumeration. Understanding P-sample nonmatches is essential to designing and evaluating coverage procedures. Any unexpected results could raise questions regarding the quality of the Accuracy and Coverage Evaluation methods and operations.

We analyzed percents not matched by age, sex, tenure, race, Hispanic Origin, and region and discussed comparable 1990 results. We also studied other operational variables to measure their effect as predictors of percents not matched and contrasted percents of P-sample not matched to corresponding statistics of E-sample erroneous enumerations, whole-person imputations, late census adds (reinstatements), and net undercount. Statistics in this study were computed using nonmovers and outmovers without adjustment for the number of inmovers, unlike official statistics which did adjust for inmovers. The differences were not expected to affect any of the significance tests or conclusions reported here.

How did the 2000 overall percent of P-sample not matched compare to 1990 results?

The percent of P-sample nonmatches was larger in 2000 than in 1990. That could be explained by the increase for 2000 in late census adds (reinstatements), whole-person imputations, and insufficient information for matching.

- The overall percent not matched for 2000 was 8.2 percent; it was statistically significantly larger than the 7.8 percent observed for 1990.

- From 1990 to 2000, the number of late census adds (reinstatements), whole-person imputations, and census records with insufficient information for matching increased. They represented census data which could not be included in matching and therefore resulted in fewer matches.

Did any unexpected result point to particular variables as predictors of percents not matched?

No, the results were consistent with 1990 results and general expectations. As expected, post-stratification variables consistently divided the sample into groups that had different percents not matched. Other results appeared consistent with historical or other expected patterns, specifically:

- Similar distributions of percents not matched for 2000 and 1990 were observed for age, sex, tenure, race, Hispanic Origin, and region.

- Other variables found to have higher percents not matched were as follows (percents not matched listed in parentheses):

  < outmovers (22.6 percent),
  < proxy respondents (20.3 percent),
  < not a single-family dwelling (13.3 percent to 22.1 percent),

- seven or more residents at the address (17.2 percent),
- distant or no kinship to person listed first on questionnaire; not parent, child, or spouse (17 percent),
- imputed post-stratification variables (13.8 percent),
- renters (13.1 percent),
- young adults, age 18 to 29 (12.2 percent),
- native Hawaiian or Pacific Islander, American Indian, Blacks, Hispanics (11.7 percent to15 percent),
- low return rates (11.1 percent).

What implications did these results have on the adjustment decision?

The apparent increase in percents of P-sample not matched from 1990 to 2000 was explained by the increase in late census adds (reinstatements) and whole person imputations for 2000. Given this, and the fact that other results were consistent with historical or other expected patterns, **no negative implications on data from the Accuracy and Coverage Evaluations were noted.**

1. BACKGROUND

This report was part of a tradition of investigations aimed at understanding P-sample nonmatches and the characteristics related to their frequency.

1.1 What was the role of P-sample nonmatches in the Census 2000 Accuracy and Coverage Evaluation (A.C.E.)?

The Accuracy and Coverage Evaluation (A.C.E.) involved two samples. Both were selected in sample areas, consisting of block clusters and sometimes a subsample within blocks. People enumerated by the census in A.C.E. sample areas made up the E sample, which was used to count errors among census enumerations. Census Day residents in sample areas listed by the A.C.E. survey made up the P sample, which was used to determine who was missed in the census. Census Day residents, if not available during the A.C.E. interview, were listed with the aid of proxy respondents. Inmovers, residents at the address during the A.C.E. interview, often were proxy respondents for outmovers, who moved out of the sample address since enumeration in the census. Names and characteristics of P sample people were compared to those of census enumerations in the sample cluster or designated surrounding blocks. Matches were persons found in both, that is, a record existed for them among both A.C.E. and Census 2000 enumerations.

The focus of this analysis was on P-sample nonmatches. It addressed only the undercount aspects of census coverage evaluation. The aim was to identify characteristics that were related to being missed in census enumeration. The statistic used in this study was the percent not matched, the percent of nonmatches among P-sample persons, computed within age, race, or other descriptive variables. The proportion not matched (NM/P), along with the number of data-defined census persons without imputation or late census adds (DD), and the proportion of correct enumerations as determined by the E sample (CE/E), had an important role in the dual system estimate (DSE) formula:

$$DSE = (DD) * (CE/E) / (1 - (NM/P)).$$

1.2 What prior reports related to these analyses?

There were many ways to investigate the nonmatches. The major approach in this report was to divide the P sample into groups on the basis of levels of important variables, compute percents not matched for those levels and test for differences. Percents not matched have been studied independently of the effects of erroneous enumerations. Erroneous enumerations in Census 2000 were investigated by Feldpausch (2001). Beaghen, Feldpausch, and Byrne (2001) modeled both E-sample and P-sample data to gain insight into missed enumerations.

Other publications provided background to this research. Hogan (1993) reported on both analyses and procedures for the 1990 census. Hogan (2000) described application of theory in A.C.E. Childers (2001) described the A.C.E. design. Adams, Barrett, and Byrne (2001) summarized procedures for A.C.E. operations.

2. METHODS

2.1 How were A.C.E. records matched to Census 2000 records?

This study used the person-level records of Census 2000 and of the independent A.C.E. enumeration. P-sample person records and census person records were computer matched within cluster. The computer matching involved first standardizing the name formats. Next, names and person characteristics of the P-sample people were compared to those of census people with sufficient information for matching and follow-up. A ranking score was assigned to each pair of person records and the optimal pairings were identified. Those pairs were reviewed to determine cutoffs in the scores taken to separate matches, possible matches, and nonmatches. Match cutoffs were assigned conservatively to avoid false matches.

The possible matches and P-sample nonmatches were clerically reviewed using an automated match and review system. The names, age, race, Hispanic origin, sex, relationship, household composition, and address were displayed for review by the matching clerks, who matched some people the computer could not. After the matching, field follow-up was conducted to confirm or resolve who matched and who should have been counted in the cluster on Census Day.

2.2 What other operations were important to the Accuracy and Coverage Evaluation?

Sampling, imputation, relisting, and targeted extended search were some of the operations conducted to generate the dual system estimates. Each record was given a sampling weight derived from the probability of selection of the block cluster. If there were eighty or more housing units in a block, a block segment was subsampled within the block and given an additional subsample weight. When housing unit follow-up showed more than eighty percent of the cluster's A.C.E. addresses were listed incorrectly, an additional listing, or relisting, of the cluster's addresses was conducted.

Sometimes census procedures assigned an address an incorrect block cluster identity code, complicating matching efforts. That error, called geocoding error, should have had no effect on census population counts and little, if any, effect on dual system estimates, but it was expected to inflate variances of the estimates. An operation, called targeted extended search (TES), was conducted to expand the area searched for geocoding errors. TES improved the precision of estimates and improved the robustness of dual system estimates. TES involved another level of sampling and weighting. Relisted clusters and other clusters with the greatest evidence of possible geocoding error were included in the operation with certainty. Other clusters with any potential of geocoding error were sampled. See Wolfgang, Stallone, and Adams (2001) for more information and results of the TES.

If match status remained unresolved, match probability was imputed. If residence status remained unresolved, residence probability was imputed. When variables used for post-stratification, namely tenure, age, sex, race, and Hispanic origin, were missing, values were imputed. For households not successfully interviewed, a non-interview adjustment was applied. Sampling weights (including weights for large block subsampling and TES selection), match probabilities, residence probabilities, and non-interview adjustments were applied in all analyses.

2.3 What statistics were analyzed?

The percent not matched, the weighted number of nonmatches divided by the weighted number of P-sample persons expressed as a percent, was computed for various groups within the P sample. Identifying groups with unusually high percents not matched provided insights into conditions associated with missed census enumerations. For this purpose, P-sample persons were grouped into meaningful levels of a variety of variables, especially variables used for post-stratification (Haines, 2001) and others that were expected to be related to the percent of nonmatches.

The P sample analyzed in this report included nonmover and outmover data. See the discussion in the Limitations section regarding how official estimates use inmover data as well. Most of the analyses were done using the whole P sample. Those analyses were conducted using variables from A.C.E. data collection or processing.

2.4 How were statistics compared?

The percents not matched of different P-sample subgroups were compared using VPLX, software designed by Fay (1990) to estimate variances in complex sample surveys using replication methods. Stratified Jackknife methods were used to compute variance estimates for the percents not matched. VPLX also generated t values for pairs of these statistics.

Statistical significance for these t values was determined using the Bonferroni multiple comparison of means technique. It controlled the probability of Type I error for a family of tests. In the context of this analysis, a family of tests was defined as all tests conducted among sample subgroups formed from the variable under analysis. For example, when comparing four subgroups, six pairs of statistics were tested. To control the chance of Type I error at $\alpha = 0.10$ for all six tests combined, we used an adjusted criterion t-value associated with the probability of one of six two-tailed tests that had a joint error probability equal to 0.10. In addition, tests with levels based on less than 100 person records were avoided, either through collapsing with other levels or simply by dropping the level from that family of tests.

3. LIMITATIONS

This analysis of A.C.E. data had certain research limits. It had a specific focus on P-sample nonmatches. It did not address the impact of other total error components (Mulry and Spencer, 1991), even errors in collected Census 2000 data measured by the E sample.

Nonmatch statistics in this analysis were different from the official statistics computed during production. Nonmatch statistics in this study were computed solely using nonmovers and outmovers; inmovers were not used. Official nonmatch statistics were computed using a combination of nonmover, outmover, and inmover information. For official dual system estimation, statistics were computed and defined for levels of post-stratum variables. In these analyses, we were interested in some non-post-stratum variables and used the simpler methodology. Haines (2001) and Davis (2001) elaborated on the different methodologies for handling movers. This analysis procedure yielded percents not matched that were a little lower than official percents not

matched, typically about 0.3 percent within major population subgroups, as seen in Davis (2001). These small, fairly consistent differences were not expected to affect any of the significance tests reported here.

Variance computations in these analyses were simplified and did not take all levels of the sampling into account. We expect only trivial impacts on variances due to variance computations; we expect no impact on test results and conclusions.

4. RESULTS

The overall percent not matched, 8.2 percent, was computed from a weighted total of 21,146,168 nonmatches and a weighted total of 258,547,382 P-sample persons. The percents reported below ranged from 5.3 percent (among spouses of the first person listed in the A.C.E. interview) to around 22.6 percent (among outmovers).

The overall percent not matched without adjusting for the number of inmovers was 8.2 percent (standard error, s.e. = 0.1 percent). The percent not matched computed with adjustment for the number of inmovers as in official dual system estimates was 8.4 percent (s.e. = 0.1 percent). Both of these Census 2000 values differed statistically from the 1990 percent, 7.8 percent (s.e. = 0.2 percent) computed using inmover data rather than outmover data.

Section 4.1 below reported analyses of percents not matched for different types of nonmatches. Nonmatch type was defined by whether others in the household were matched to census persons and by whether the housing unit address was matched to a census address. Section 4.2 reported analyses of subgroups of the P sample showing what characteristics were related to high percents not matched. Data from 1990 were discussed when available for the most important variables.

4.1 How did nonmatch type among nonmovers relate to percent not matched?

In 1990, the nonmover P-sample nonmatches that were resolved (without imputation of residence or match status) were classified by match status of the whole household into three categories: partial-household nonmatches, whole-household nonmatches in matched housing units, and whole-household nonmatches in nonmatched housing units. The 1990 data (Hogan, 1993) and the 2000 data were presented as a percent not matched out of the total of resolved P-sample nonmovers. Thus, the percents added up to overall resolved nonmover percents not matched of 7.2 percent and 5.9 percent.

Nonmatch Household Types: Percent of Resolved Nonmovers Not Matched for 2000 and 1990 (Table 4.1.1)

Nonmatched Person by Household Type	2000		1990	
	Percent of Nonmovers	Percent of Nonmatches	Percent of Nonmovers	Percent of Nonmatches
Partial-Household Nonmatch	2.2	30.0	1.8	30.4
Whole-Household Nonmatch in a Matched Housing Unit	3.3	45.9	2.3	38.5
Whole-Household Nonmatch in a Nonmatched Housing Unit	1.7	24.1	1.8	31.1
TOTAL	**7.2**	**100**	**5.9**	**100**

Note: Movers and unresolved cases were removed from both 1990 and 2000 analyses.
1990 results come from H. Hogan (1993).
Five 1990 categories were combined for this report into these three categories for the sake of comparing 1990 and 2000 analyses (nonmatched persons living in a building missed by the census were collapsed into whole-household nonmatches in nonmatched housing units; persons listed on a census form returned but not processed were collapsed into whole-household nonmatches in matched housing units).

The percent of nonmatches among nonmovers increased from 1990 to 2000. Increased housing unit matches may have been related to that increase. Stronger explanations, however, come from trends in late census adds (reinstatements), whole-person imputations and person records with insufficient information for matching and follow-up. See Table 4.1.2. Records for these enumerations were not available for matching and therefore became nonmatches. In particular, proportionately more of the nonmatches were in matched housing units. The sharper increase in percents for whole-household nonmatches in matched housing units coincided with increases in whole-household imputations and late census adds (reinstatements). See Appendix A for late census add and imputation data for age/sex, tenure, and race/Hispanic origin domain groups.

For whole-household nonmatches in matched housing units, the most prevalent census statuses were: whole-household insufficient information for matching and follow-up, conflicting households, vacant housing units, and housing units with no data-defined people. Table 4.1.3 lists these results.

Number of Records Not Available for Matching in 2000 and 1990 (Table 4.1.2)

Type of Records Not Available	in 1990	in 2000
Whole-Person Imputations: Imputed Pop Count	53,655	1,172,144
Whole-Person Imputations: Other Whole-Household Imputations	1,547,101	2,269,010
Whole-Person Imputations: Other Whole-Person Substitutions	300,652	2,333,112
Late Census Adds (Reinstatements)	Not Available	2,239,427
Insufficient Information for Matching and Follow-up	2,859,159	4,781,418

Percent Not Matched for Census Status of Whole-Household Resolved P-Sample Nonmover Nonmatches in Matched Housing Units (Table 4.1.3)

Status of Matched Census Unit	Weighted P-Sample Nonmovers	Percent of Total	Percent of Non-movers	Non-Data-Defined People in Matched Census Units	E-sample People in the Matched Census Units	Total Census People in Matched Census Units
Whole-household Insufficient Information for Matching and Follow-up	1,934,935	23.74	0.78	164,264	1,780,076	1,944,340
Conflicting Households *	1,892,421	23.22	0.77	65,876	2,436,403	2,502,279
Vacant Housing Units	1,530,131	18.77	0.62	Not Available	0	0
Housing Units with No Data-Defined People	1,528,553	18.75	0.62	1,586,979	0	1,586,979
At Least One Person Matched or Possibly Matched to Another P-Sample Housing Unit	691,030	8.48	0.28	13,689	815,210	828,899
Housing Units Matched to a Surrounding Block	569,204	6.98	0.23	Not Available	Not Available	Not Available
At Least One Person Matched or Possibly Matched to the P-Sample Housing Unit **	4,565	0.06	0.00	0	2,997	2,997
Total	**8,150,839**	**100.00**	**3.31**	**1,900,224**	**5,832,214**	**7,732,438**

Note: Late census adds (reinstatements) were not among these persons; their housing units were not available for final housing unit matching.

* Whole households of non-E-sample people with sufficient information for matching and follow-up were included with the conflicting households.

** Whole-household nonmatches include housing units that contain a mix of possible matches

and nonmatches. If the possible match was at the address match, the people appeared here.

4.2 What P-sample variables related to percent not matched over the general sample?

This section provides results for relating other characteristics of clusters, households, or persons to percent not matched. Results are presented in tables displaying variable level names, percents not matched (in the column headed "Percent"), the rank of that level's percent not matched from lowest to highest ("Rank"), a list of the rank numbers with which a significant difference was found ("Differs from"), the stratified jackknife standard error ("s.e."), and the weighted percent of persons comprising that level's subgroup ("n"). Criterion t-values (such as $| t | > 1.65$) vary, as described above, with the number of comparisons being made in the family of tests. The criterion that applies in that table was noted below each table.

Descriptions and interpretations of results accompany the main tables. Other tables with alternative or related results are placed in Appendix B.

Important variables were grouped in several categories:

- Variables used in defining post-strata for dual system estimation
- Other nation-wide variables relevant to sampling or estimation
- Other household-level variables
- Other person-level variables

4.2.1 What post-stratification variables related to percent not matched?

Variables used to form post-strata in dual system estimation (Haines, 2001) were of primary interest. They were analyzed here using the levels as defined for post-stratification. Levels of MSA/TEA, Return Rate, and Region were used in some but not all post-stratum group definitions. All P-sample persons were included in these analyses, regardless of whether their post-stratum was affected by the variable.

Age/Sex, Tenure, Race/Hispanic origin domain, and Region tables below include 1990 results reported in the Analysis of Movers report (Liu, Byrne, and Imel, 2001). The 2000 data were based on nonmovers and outmovers without adjustment for the number of inmovers. In contrast, the 1990 data were based on nonmovers and inmovers, as the official estimates were computed in that census.

Percents not matched differed by age and sex post-strata levels except in one comparison involving children. Generally, from young adults to the elderly, the percent not matched decreased, with males generally having higher percents not matched at each age. Children's percents not matched were close to the median of groups aged 18 to 49, commonly child-raising ages. Results based on age without gender breakdowns (See Table B1 in Appendix B) showed the same. Percents not matched were related to the age of adults. We might also speculate that the child's percents not matched relate to their parents' ages. Ignoring age, males had a percent of 8.8 percent compared to 7.6 percent for females (See Table B2 in Appendix B).

Percent Not Matched by Age and Sex (Table 4.2.1)

	2000					1990	
Age/Sex	**Percent**	**Rank**	**Differs from**	**s.e.**	**n**	**Percent**	**s.e.**
0-17	8.8	5	1,2,3,6,7	0.2	26.2	8.5	0.3
18-29 Male	13.2	7	all	0.3	7.5	13.3	0.4
18-29 Female	11.1	6	all	0.2	7.7	11.6	0.3
30-49 Male	8.5	4	1,2,3,6,7	0.2	15.2	7.9	0.3
30-49 Female	6.9	3	all	0.1	16.2	6.2	0.2
50+ Male	6.2	2	all	0.2	12.3	4.8	0.2
50+ Female	5.6	1	all	0.1	14.9	4.1	0.2

Note: Criterion for levels reported to differ was $|t| > 2.815$

Typically, home owners have had a much lower percent not matched than non-owners.

Percent Not Matched by Home Ownership (Table 4.2.2)

	2000					1990	
Tenure	**Percent**	**Rank**	**Differs from**	**s.e.**	**n**	**Percent**	**s.e.**
Owner	6.1	1	2	0.1	69.8	5.1	0.2
Non-owner	13.1	2	1	0.2	30.2	13.6	0.4

Note: Criterion for levels reported to differ was $|t| > 1.645$

Respondents to Census 2000 were able to self-identify with more than one race group. Combining 63 levels of Race with two levels of Hispanic Origin yielded 126 possible Race/Hispanic Origin groups. Rules were adopted to assign persons in those 126 groups to one of seven Race/Hispanic Origin Domains (See Haines, 2001).

Hispanic, Non-Hispanic Black, American Indian on Reservation, American Indian off Reservation, and Native Hawaiians or Pacific Islander Domains had higher percents not matched than the Non-Hispanic White or "Some other race" Domain. The Non-Hispanic Asian Domain had a higher percent not matched than the Non-Hispanic White or "Some other race" Domain and had a lower percent than three other levels. Other differences were not significant. Caution should be used in comparing 1990 to 2000 results because differences in definitions for 1990 race/ethnicity categories were not the same as for 2000 race and Hispanic origin categories.

Percent Not Matched by Race/Hispanic Origin Domain (Table 4.2.3)

2000 Race/Hispanic Origin Domain	Percent	Rank	Differs from	s.e.	n
American Indian on Reservation	13.7	6	1,2	1.1	0.2
American Indian off Reservation	11.7	3	1	1.1	0.5
Hispanic	12.1	4	1,2	0.3	12.3
Non-Hispanic Black	12.8	5	1,2	0.3	11.4
Native Hawaiian or Pacific Islander	15.0	7	1	2.5	0.2
Non-Hispanic Asian	9.2	2	1,4,5,6	0.5	3.4
Non-Hispanic White or "Some Other Race"	6.7	1	all	0.1	72.1
TOTAL for 2000	**8.2**				**100.0**

1990 Race / Ethnicity	Percent	s.e.
American Indian on Reservation	21.9	1.4
Non-Black, Non-Asian, Non-American-Indian Hispanic	12.6	0.6
Black	14.2	0.5
Asian or Pacific Islander	9.8	0.9
Non-Hispanic White / Other	6.2	0.2
TOTAL for 1990	**7.8**	

Note: Criterion for levels reported to differ was $|t| > 2.815$

Metropolitan Statistical Area (MSA) and Type of Enumeration Area (TEA) were combined to form one variable used in post-stratification. MSAs denoted the boundaries of cities or other areas named for statistical purposes. Most of the population was in the Mailout/Mailback TEA, in which people receive and return census forms by mail. Mailout/Mailback areas were divided into three levels based on size of the MSA. A fourth level was comprised of other areas where census workers visited to list or update addresses or conduct enumerations on the spot.

The extremes, large MSAs and areas where enumeration was not conducted by mail, had higher percent not matched. Perhaps the most urban and the most rural areas have different causes (possibly mobility for large MSAs and inaccessibility for very rural areas) for being harder to enumerate than the more developed rural and suburban areas that had unique postal addresses.

A complete analysis of seven types of enumeration level (See Appendix B, Table B3) showed few significant differences because most of the levels had large standard errors and thus were not very precise or discriminating in defining groups with different percents not matched.

Percent Not Matched by Size of Metropolitan Statistical Area and Type of Enumeration Area (Table 4.2.4)

MSA/TEA	Percent	Rank	Differs from	s.e.	n
Large MSA, Mailout/Mailback	9.0	3	1,2	0.2	30.4
Medium MSA, Mailout/Mailback	7.4	2	3,4	0.2	31.3
Small MSA & Non-MSA Mailout/Mailback	7.3	1	3,4	0.2	20.2
All Other TEAs	9.2	4	1,2	0.3	18.1

Note: Criterion for levels to differ was $|t| > 2.386$

The tract-level return rate, a sign of public cooperation, was the proportion of occupied housing units in a census tract that returned a 2000 census questionnaire. High and low return rate indicator values were assigned for the Non-Hispanic White or "Some other race," Non-Hispanic Black, and Hispanic domains. Persons in all other Race/Hispanic Origin Domains were assigned a return rate indicator value of "Not Applicable" since they were not post-stratified by return rate (Haines, 2001). Persons in high return rate post-strata had a lower percent not matched than other P-sample persons.

Percent Not Matched by Return Rate Indicator (Table 4.2.5)

Return Rate Indicator	Percent	Rank	Differs from	s.e.	n
High	7.1	1	all	0.1	72.3
Low	11.1	3	all	0.3	23.4
Not Applicable	10.0	2	all	0.5	4.3

Note: Criterion for levels to differ was $|t| > 2.121$

Another national variable of interest was region of the U.S. The Midwest region stood out with a lower percent not matched than other areas.

Percent Not Matched by Region of the United States (Table 4.2.6)

	2000					1990	
Region	**Percent**	**Rank**	**Differs from**	**s.e.**	**n**	**Percent**	**s.e.**
Northeast	8.3	2	1	0.3	19.0	8.2	0.5
Midwest	6.1	1	all	0.2	22.9	5.7	0.3
South	9.1	4	1	0.2	35.3	8.6	0.3
West	8.7	3	1	0.3	22.8	8.3	0.3
TOTAL	8.2				100.0	7.8	

Note: Criterion for levels to differ was | t | > 2.386

4.2.2 What other estimation variables related to percent not matched?

Age, sex, race, Hispanic origin, and tenure were sometimes imputed for the A.C.E. Where information was complete enough to require no imputation of post-stratum characteristics, percents not matched were lower.

Percent Not Matched by Imputation of Characteristics (Table 4.2.7)

Imputed or Not	**Percent**	**Rank**	**Differs from**	**s.e.**	**n**
Not Imputed	7.9	1	2	0.1	94.7
Imputed	13.8	2	1	0.4	5.3

Note: Criterion for levels to differ was | t | > 1.645

Subsampling of housing units within clusters with a large number of housing units was done to reduce the intra-cluster correlation and to reduce the interviewing workloads. Subsampled clusters had a higher percent not matched.

Percent Not Matched by Subsampling Involvement (Table 4.2.8)

Subsampled or Not	**Percent**	**Rank**	**Differs from**	**s.e.**	**n**
Not Subsampled	7.8	1	2	0.1	63.5
Subsampled	8.8	2	1	0.2	36.5

Note: Criterion for levels to differ was | t | > 1.645

4.2.3 What household-level variables related to percent not matched?

Some characteristics of the household appeared likely to relate to percent not matched, including the type of structure at the basic address, household size, and whether the respondent was a proxy or household member.

A significantly lower percent of nonmatches was found among residents of single family permanent structures relative to mobile homes, multiunit structures (like apartment buildings), or miscellaneous other structures (like living quarters within a special place). See Appendix B Table B4 for an additional analysis that combines mobile home with Living Quarters in a Special Place and Unclassified, leading to the same conclusion.

Percent Not Matched by Type of Structure at Basic Address (Table 4.2.9)

Structure	Percent	Rank	Differs from	s.e.	n
Single-Family Dwelling	6.5	1	all	0.1	75.7
Multi-Unit	13.4	3	1	0.3	18.9
Mobile Home	13.3	2	1	0.7	5.2
Living Quarters in a Special Place and Unclassified	22.1	4	1	6.0	0.2

Note: Criterion for levels to differ was $|t| > 2.386$

Household size was based on the number of P-sample persons enumerated at the address. Up to six persons could be enumerated on most census forms; additional forms or procedures were usually needed to enumerate seven or more. Large households (perhaps those with more residents than would fit on a census questionnaire without a continuation form) had a higher percent not matched. Households with only one reported resident had a higher percent not matched than those with a few other residents. If solitary residents were more likely to be mobile, mobility could be related to percent not matched.

Percent Not Matched by Household Size (Table 4.2.10)

Household Size	Percent	Rank	Differs from	s.e.	n
One person	10.5	2	all	0.2	10.4
2-6 persons	7.4	1	all	0.1	84.9
7 or more persons	17.2	3	all	0.5	4.7

Note: Criterion for levels to differ was $|t| > 2.121$

When an interview could not be obtained from someone in the household, a proxy may have provided the data. As expected, when the respondent was a member of the household and not a proxy, the percent not matched was lower.

Percent Not Matched by Type of Respondent (Table 4.2.11)

Type of Respondent	Percent	Rank	Differs from	s.e.	n
Household Member	7.5	1	2	0.1	94.5
Proxy	20.3	2	1	0.4	5.5

Note: Criterion for levels to differ was | t | > 1.645

4.2.4 What person-level variables related to percent not matched?

One variable based on person-level data involved movers. Since nonmovers were usually household members, a combined analysis was conducted. Household member respondents who were not movers had a low percent not matched, while proxy nonmovers had a high percent not matched and both outmover groups had even higher percents not matched. The household member, outmover group was comprised of movers for whom a nonmover reported. That showed that while proxy related to percent not matched without the influence of movers, the mover variable had the stronger relationship to percent not matched. See Appendix B, Table B5 for the corresponding 2000 analysis not involving proxy, which also showed outmovers had the higher percent not matched.

Percent Not Matched by Proxy and Mover Status (Table 4.2.12)

Proxy and Mover Status	Percent	Rank	Differs from	s.e.	n
Household Member, Nonmover	7.3	1	all	0.1	93.7
Household Member, Outmover	21.5	3	1,2	0.8	0.8
Proxy, Nonmover	18.1	2	all	0.5	2.9
Proxy, Outmover	22.9	4	1,2	0.6	2.6

Note: Criterion for levels to differ was | t | >2.386

Another interesting analysis involved Kinship, a measure of how closely related the person was to the central person in the household. The respondent for A.C.E. 2000 designated someone in the household, usually the person in whose name the residence was owned or rented, to be listed first in the data collection. The question, "What is . . . 's relationship to . . . ?" was asked about subsequent persons listed and referred to the first person listed. Kinship for other persons in the household was in relationship to this reference person. Reference persons either lived alone or shared the housing unit with other relatives or nonrelatives.

The less closely related a person was to the reference person, the higher the percent not matched. Reference persons living alone had a higher percent not matched than those living with others. A spouse's percent not matched was the lowest. Perhaps these kinship categories reflected mobility or how likely household members were to move.

Percent Not Matched by Kinship to Reference Person (Table 4.2.13)

Kinship	Percent	Rank	Differs from	s.e.	n
Reference person, alone	10.2	4	all	0.2	10.1
Reference person, not alone	6.2	2	all	0.1	28.4
Spouse	5.3	1	all	0.1	20.2
Parent / Child	8.0	3	all	0.1	30.4
Other relatives and nonrelatives	17.0	5	all	0.3	10.9

Note: Criterion for levels to differ was $|t| > 2.568$

REFERENCES

Adams, T. (2001), Barrett, D., and Byrne, R.. "Operational Plan for Accuracy and Coverage Evaluation (A.C.E.) for Census 2000," DSSD Census 2000 Procedures and Operations Memorandum Series S-TL-06, U.S. Census Bureau, Washington, D.C.

Beaghen, M., Feldpausch, R., and Byrne, R. (2001). "Modeling Accuracy and Coverage Evaluation Non-matches in the Census 2000," Proceedings of the Section on Survey Research Methods, American Statistical Association, to appear.

Childers, D. (2001). "The Design of the Census 2000 Accuracy and Coverage Evaluation (A.C.E.)" DSSD Census 2000 Procedures and Operations Memorandum Series S-DT-01, U.S. Census Bureau, Washington, D.C..

Davis, P. (2001). "Accuracy and Coverage Evaluation: Dual System Estimation Results," DSSD Census 2000 Procedures and Operations Memorandum Series B-9*, U.S. Census Bureau, Washington, D.C.

Fay , R. (1990). "VPLX: Variance Estimates for Complex Samples," Proceedings of the Section on Survey Research Methods, American Statistical Association, 266-271.

Feldpausch, R. (2001). "Census 2000 E-Sample Erroneous Enumerations," Proceedings of the Section on Survey Research Methods, American Statistical Association, to appear.

Hogan, H. (1993). "The 1990 Post-Enumeration Survey: Operations and Results," Journal of the American Statistical Association, 88, 1047-1060.

Hogan, H. (2000). "Accuracy and Coverage Evaluation: Theory and Application," Internal document, U.S. Census Bureau, Washington, D.C.

Haines, D. (2001). "Accuracy and Coverage Evaluation Survey: Computer Specifications for Person Dual System Estimation (U.S.) -Re-issue of Q-37," DSSD Census 2000 Procedures and Operations Memorandum Series Q-48, U.S. Census Bureau, Washington, D.C.

Liu, X., Byrne, R., and Imel, L. (2001). "Executive Steering Committee on Accuracy and Coverage Evaluation Policy II Report Number 15: Analysis of Movers," DSSD Census 2000 Procedures and Operations Memorandum Series T-10, U.S. Census Bureau, Washington, D.C.

Mulry, M. and Spencer, B. (1991). "Total Error in PES Estimates of Population," Journal of the American Statistical Association, 86, 839-854.

Wolfgang, G., Stallone, P., and Adams, T. (2001). "Targeted Extended Search in the Accuracy and Coverage Evaluation of the Census 2000," Proceedings of the Section on Survey Research Methods, American Statistical Association, to appear.

APPENDIX A

CENSUS AND CENSUS COVERAGE STATISTIC TABLES

These tables provided comparison of percents of P-sample nonmatches, E-sample erroneous enumerations, whole-person imputations, late census adds (reinstatements), and net undercount for Age/sex, Tenure, and Race/Hispanic Origin Domain. The P-sample nonmatch percentages were weighted numbers of nonmatches among nonmovers and outmovers, without adjusting for the number of inmovers. The percent of erroneous enumerations was weighted with the final E-sample weights. The late census adds (reinstatements) were the people added back after the potential duplicates were reinstated, which included late census adds that were whole-person imputations.

Census and Census Coverage Statistics for Age and Sex (Table A1)

	Percent P-Sample Nonmatch	Percent E-sample Erroneous Enumeration	Percent Whole-Person Imputation	Percent Late Census Adds	Percent Net Under-count
0-17	8.8	4.1	3.1	0.9	1.5
18-29 Male	13.2	7.1	2.9	0.8	3.8
18-29 Female	11.1	6.4	2.6	0.9	2.2
30-49 Male	8.5	4.8	1.8	0.8	1.9
30-49 Female	6.9	4.0	1.6	0.8	1.0
50+ Male	6.2	4.7	1.2	0.8	-0.2
50+ Female	5.6	4.5	1.3	0.8	-0.8
Total	**8.2**	**4.7**	**2.1**	**0.8**	**1.2**

Census and Census Coverage Statistics for Tenure (Table A2)

	Percent P-Sample Nonmatch	Percent E-sample Erroneous Enumeration	Percent Whole Person Imputation	Percent Late Census Adds	Percent Net Under-count
Owner	6.1	3.6	1.6	0.8	0.4
Non-Owner	13.1	7.3	3.1	1.1	2.8
Total	**8.2**	**4.7**	**2.1**	**0.8**	**1.2**

Census and Census Coverage Statistics for Race and Hispanic Origin (Table A3)

	Percent P-Sample Nonmatch	Percent E-sample Erroneous Enumeration	Percent Whole Person Imputation	Percent Late Census Adds	Percent Net Under-count
American Indian on reservation	13.7	4.2	5.0	1.0	4.7
American Indian off reservation	11.7	6.0	2.6	1.2	3.3
Hispanic	12.1	5.5	3.9	1.0	2.8
Non-Hispanic black	12.8	7.3	3.4	1.0	2.2
Hawaiian or Pacific Islander	15.0	7.0	3.7	0.9	4.6
Non-Hispanic Asian	9.2	5.4	2.8	0.7	1.0
Non-Hispanic white	6.7	4.1	1.5	0.8	0.7
Total	**8.2**	**4.7**	**2.1**	**0.8**	**1.2**

APPENDIX B

The following tables elaborated or showed a different view of the main analyses.

Percent Not Matched by Age – Refer to Table 4.2.1 (Table B1)

Age	Percent	Rank	Differs from	s.e.	n
0-17	8.8	3	all	0.2	26.2
18-29	12.2	4	all	0.2	15.1
30-49	7.7	2	all	0.1	31.4
50+	5.9	1	all	0.1	27.3

Note: Criterion for levels to differ was | t | > 2.386

Percent Not Matched by Sex – Refer to Table 4.2.1 (Table B2)

Sex	Percent	Rank	Differs from	s.e.	n
Male	8.8	2	all	0.1	48.1
Female	7.6	1	all	0.1	51.9

Note: Criterion for levels to differ was | t | > 1.645

Percent Not Matched by TEA – Refer to Table 4.2.4 (Table B3)

Type of Enumeration Area	Percent	Rank	Differs from	s.e.	n
Mailout / Mailback	8.0	2	6	0.1	81.9
Update / Leave	8.8	3	6	0.4	16.9
List / Enumerate	17.8	7	none	3.9	0.3
Rural Update / Enumerate	14.4	6	2, 3	1.5	0.5
Urban Update / Leave	9.7	4	none	3.6	0.1
Urban Update / Enumerate	7.9	1	none	4.0	0.1
Mailout / Mailback Converted to Update / Leave	13.5	5	none	4.4	0.2

Note: Criterion for levels to differ was | t | > 2.815

Percent Not Matched by Type of Structure at Basic Address – Refer to Table 4.2.9 (Table B4)

Type of Structure	Percent	Rank	Differs from	s.e.	n
Single-Family Dwelling	6.5	1	all	0.1	75.7
Multi-Unit	13.4	2	1	0.3	18.9
Mobile Homes, Living Quarters in a Special Place and Unclassified	13.6	3	1	0.7	5.4

Note: Criterion for levels to differ was | t | > 2.121

Percent Not Matched by Mover Status – Refer to Table 4.2.12 (Table B5)

Person Mover Flag	Percent	Rank	Differs from	s.e.	n
Nonmover	7.7	1	2	0.1	96.6
Outmover	22.6	2	1	0.5	3.4

Note: Criterion for levels to differ was | t | > 1.645